Predicting Solar Max DC Power

Using a Linear Regression Model

APPROVED BY

SUPERVISING COMMITTEE:

W.Mack Grady, Supervisor

Surya Santoso

Predicting Solar Max DC Power Using a Linear Regression Model

by

Youngsung Kwon, B.S.E.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN ENGINEERING

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2012

Dedicated to my lovely wife Raehee Park.

Acknowledgments

I would like to express my deep gratitude to my advisor, Dr. W. Mack Grady, for his guidance, encouragement, and financial assistance. I admire his passionate pursuit of knowledge.

I thank my fellow office mates, Joonhyun Kim, Rossen Tzartzev and Moses Kai for providing excellent advice.

Last but not least, I would like to thank my family members, especially my lovely wife, Raehee Park, and my parents for their boundless support and belief in me during my graduate studies.

Predicting Solar Max DC Power

Using a Linear Regression Model

Youngsung Kwon, M.S.E.
The University of Texas at Austin, 2012

Supervisor: W.Mack Grady

The increase in the consumption of energy year after year emphasizes the importance of power production by photovoltaic (PV) systems. Despite an increase in the use of PV systems, accurate solar power $[kWh]$ daily harvest predicting data are not readily available. Accurate predicted solar power data is necessary because the data is helpful to designers who need to optimally size a PV panel before installation. Moreover, accurately predicted max dc power can indicate whether the PV panel is operating efficiently and economically or not. This thesis develops an approach to predict max solar power based on a Linear Regression model. The approach, which ia a simple regression was implemented using measured data on a response variable, a max solar power ($P_{\max}$), and predictor variables such as Global Horizontal (GH), Plane of Array (PA), Short Circuit Current (I_{sc}), Open Circuit Voltage (V_{oc}), and Panel Temperature (T_{emp}). The statistical results of the linear regression model produced reasonable values which agreed with those of the measured data from the solar panel.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Depletion of fossil fuel reserves, significant increases in fuel prices, and global warming are some of the reasons that modern society is focusing on energy solutions. Renewable energy, especially solar power has received considerable attention as one solution to the growing energy challenge. Electric energy is produced through PV systems by converting incident solar radiation. This process has a lower environmental impact when compared to conventional energy technology. Moreover, solar power is highly abundant compared to conventional energy sources, which will some day be depleted.

Solar power production from solar radiation requires solar power data that is modeled accurately for designers who need to optimally size a PV panel before installation. Moreover, accurately predicted daily energy harvest are helpful to PV panel users to indicate whether the PV panel is operating efficiently and economically or not.

This thesis develops an predicting scheme to derive more precise solar power data using a statistical method, a simple linear regression model. This linear regression model is based on measured radiation data, GH and PA, and on the other measured PV systems data, I_{sc}, V_{oc}, T_{emp} and P_{max}. The P_{max} values and the other measured data have mathematical functions with high correlations. The relationships between measured P_{max} and the others show how the P_{max} values change according to the values of the other variables.

The object of this thesis is to develop a mathematical approach using solar data variables to obtain the most precise predicted solar power data for PV systems.

Chapter 2

Simple Linear Regression Analysis

In order to predict the max solar power, a statistical method, linear regression model can be used. A linear regression model by the least-squares method is a way of fitting a straight line model to observed data. Namely, it quantifies how a response variable is related to a set of predictor variables. If the parameters are in linear models, estimation is based on methods from linear algebra that minimize the residuals (errors). This section elaborates on the procedure of developing linear regression model.

2.1 Simple Linear Regression Model

Here a response variable y is modeled as a combination of predictor variable x, constant β_0, β_1 and error ε.

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2.1}$$

This model is probabilistic, because the uncontrolled factors are modeled by ε. The error term ε is a random variable and it is assumed to be uncorrelated and distributed with mean 0 (constant variance).

2.1.1 The Method of Least Squares

The method of least square chooses the regression coefficients β_0 and β_1 so as to minimize the sums of square shown below,

$$SS = \sum_{i=1}^{n} \varepsilon_i^2 \tag{2.2}$$

where n=the number of observations. Equation (2.1) is given by

$$\varepsilon = y - \beta_0 - \beta_1 x.$$

The minimum value of SS can be found by taking the derivative, $\dfrac{\partial SS}{\partial \beta}$,

$$\frac{\partial SS}{\partial \beta_0} = -2\sum_{i=1}^{n} \varepsilon_i = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0,$$

$$y_i - \beta_0 - \beta_1 x_i = 0 \tag{2.3}$$

$$\frac{\partial SS}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i\varepsilon_i = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$x_i(y_i - \beta_0 - \beta_1 x_i) = 0. \tag{2.4}$$

By dividing the above two expressions (2.3) and (2.4) by n, both equations becom

$$\beta_0 + \beta_1 \bar{x} = \bar{y} \tag{2.5}$$

$$\beta_0 \bar{x} + \beta_1 \overline{x^2} = \overline{xy} \tag{2.6}$$

where $\bar{x}$ is $\frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y}$ is $\frac{1}{n}\sum_{i=1}^{n} y_i$ which are called sample mean. Subtract $\beta_1\bar{x}$ on both side of Equation (2.5) and then substitute β_0 into Equation (2.6)

$$\beta_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - \bar{x}\cdot\bar{x}}. \tag{2.7}$$

The sample variance S_{xx} is defined

$$\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2.8}$$

and then

$$S_{xx} = \overline{(x-\bar{x})^2} = \overline{x^2 - 2x\cdot\bar{x} + \bar{x}^2} = \overline{(x^2)} - (\bar{x})^2 \tag{2.9}$$

since $\overline{x\cdot\overline{x}} = (\overline{x})^2$. As shown above, the sample covariance is defined by average of $(x - \overline{x})(y-\overline{y})$. Therefore,

$$\begin{aligned} S_{xy} &= \overline{(x-\bar{x})(y-\overline{y})} = \overline{xy - x\cdot\overline{y} - \bar{x}\cdot y + \bar{x}\cdot\overline{y}} \\ &= \overline{(xy)} - \bar{x}\cdot\bar{y} - \bar{x}\cdot\bar{y} + \bar{x}\cdot\bar{y} = \overline{(xy)} - \bar{x}\cdot\bar{y} \end{aligned} \tag{2.10}$$

From Equation (2.9) and (2.10), β_0 and β_1 lead to the least squares estimates

$$\beta_0 = \bar{y} - \beta_1\bar{x} \tag{2.11}$$

$$\beta_1 = \frac{S_{xy}}{S_{ss}} \tag{2.12}$$

where Equation (2.12) is called the sample correlation coefficient [1].

2.1.2 Sum of Squares (SS)

In order to study the performance of sum of squares several terms are used:

1) Observed value y_i,

2) Fitted value $\hat{y}_i$,

3) Mean value $\bar{x}, \bar{y}$.

The variation in the responsible variable y can be accounted by according to the variation in the predictor variables. The variation of the observation y_i around their mean $\bar{y}$ is defined by

$$SSTO = \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{2.13}$$

Equation (2.13) shown above refers to as the total of squares and denote it by SSTO. This SSTO can be divided into two parts

$$SSTO = SSR + SSE$$

because $SSTO$ is

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$

$$= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{2.14}$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{2.15}$$

One portion, Equation (2.14) is called the regression sum of squares, SSR. SSR measures the variation between the fitted values $\hat{y}_i$ and the mean value $\bar{y}$. If the fitted regression line passes through all of the observations, then the model explains all of the variability of the observations. Therefore, the model sum of squares equals the total sum of squares. Figure 2.1(a) indicates $SSR = SSTO$.

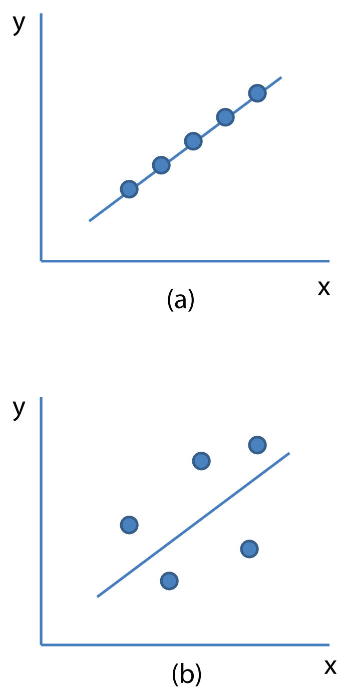

Figure 2.1: A regression model with no variation (a) and a regression model with some variation (b)

The variation Figure 2.2(a) can be described by the model but the variation in (b) is unexplained. Since this portion of the total variability is not explained by the model,the variation is called the residual sum of squares or the error sum of squares, SSE, as shown above Equation (2.15). SSE measures the variation between y_i and the fitted values $\hat{y}_i$. Moreover, these two SSR and SSE can be referred to as the analysis of variance on regression model. The deviations for the three sum of squares are shown in Figure 2.2

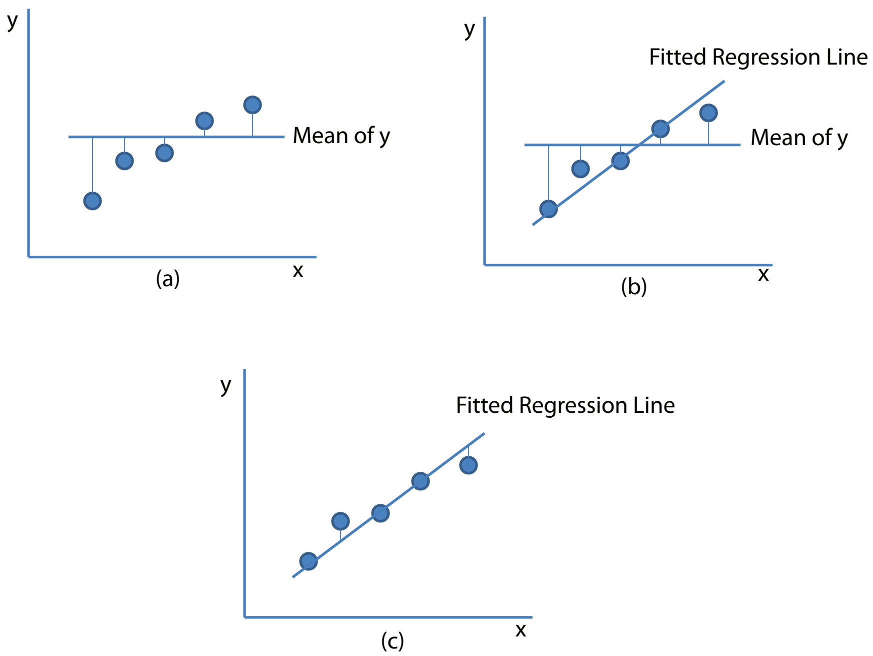

Figure 2.2: Deviations for the sum of squares. Deviations for $SSTO$(a), SSR(b), and SSE (c)

In addition, the coefficient of determination, R^2 can be a criteria for measuring the amount of variability since the coefficient of determination is the ratio of the regression sum of squares to the total sum of squares. R^2 can take on values between 0 and 1 since $SSTO = SSR + SSE$.

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{2.16}$$

2.1.3 Mean Squares

Mean squares are obtained by dividing the sum of squares by the respective degrees of freedom (p is number of predictor variables). As shown equation below, the error mean square, MSE, can be obtained as

$$MSE = \frac{SSE}{n-p-1}. \tag{2.17}$$

Similarly, the regression mean square, MSR, can be obtained by dividing the regression sum of squares by the respective degrees of freedom as follows [2]

$$MSR = \frac{SSR}{p}. \tag{2.18}$$

Chapter 3

Data and a Linear Regression Model Diagnostic Checking

3.1 Data

The data used in this study is measured data from PV systems located at The University of Texas at Austin. The measured data comprising the following components was recorded at 1-minute intervals during the whole month of January, February and March 1st through 18th, 2012:

1) PV Max DC Power ($p_{\max}$): A measured max solar power data from PV systems.
2) Global Horizontal (GH): GH is the total solar radiation from the entire sky on a horizontal surface which includes the sum of the direct-beam, diffuse and reflected solar radiation [3]. Typically, 1-minute interval or 5-minute interval GH data are available and agriculture stations typically use hourly average data; however, our GH data are recorded every 5-second.
3) Plane of Array (PA): PA is identical with GH; however, it is tilted thirty degrees from GH to the direction of the solar panel.
4) Short Circuit Current (I_{sc}): I_{sc} in Figure 3.1 flows through the short circuit when the solar cell is short-circuited and then no current flows through the diode [4].
5) Open Circuit Voltage (V_{oc}): V_{oc} is produced when the solar cell is open circuited and

then all of the I_{sc} flows through the diode.

6) Panel Temperature (T_{emp}): The solar panel temperature.

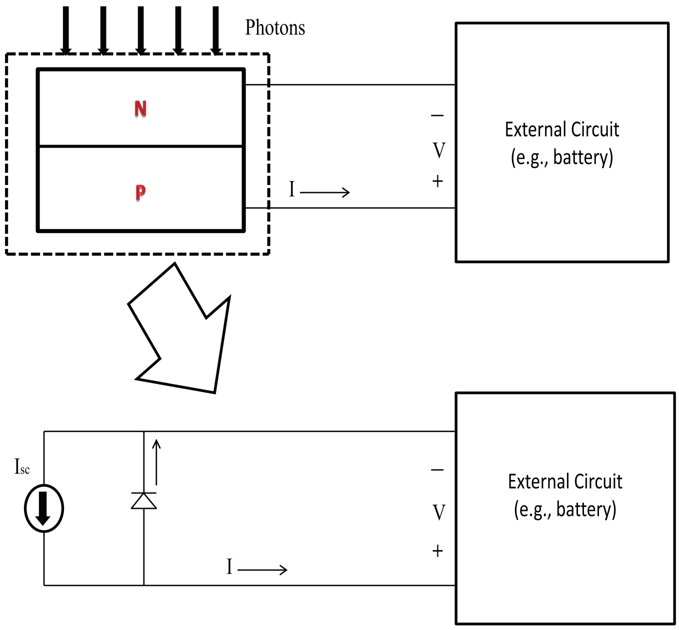

Figure 3.1: Equivalent Circuit of a PV Solar Cell for I_{sc}

3.2 Diagnostic Checking

The scatter plots in the next chapter, PV Max DC Power versus Global Horizontal, Plane of Array, and Short Circuit Current consecutively show that the correlation of each response variable and predictor variable is approximately linear, except for x=Open Circuit Voltage and Panel Temperature which suggests a linear regression model,

$$\hat{Y} = b_0 + b_1 X. \tag{3.1}$$

When building an estimating linear model such as the equation shown above, it is important to consider the ANOVA table which is used to evaluate the suitability of the linear regression model.

3.2.1 ANOVA: Analysis of Variance

The analysis of variance, ANOVA, is a method to test for the significance of fitted regression model. As the name implies, this approach uses the variance of the observed data to fitted regression line if a regression model can be applied to the observed data. The observed variance is partitioned into components which are used in the test for significance on the fitted regression line.

The ANOVA table is comprised of several columns labeled Sum of Squares (SS) , Degrees of Freedom (df) , Mean Square (MS) , F test and p-value. There is no indicated significance level α which indicates error probability in the ANOVA table but we always assume that α is 0.05. Of all the information presented in the ANOVA table, the p-value is major interest which indicates the probability assuming that there is no linear association in regression model (null hypothesis, H_0). If a significant test gives a p-value lower than the α, the null hypothesis, H_0, is rejected and this result can be referred to as statistically significant. Moreover, the F test can also be a criterion of significance of fitting model. If a high value of F test has a lower p-value, the result indicates that the fitted regression model is significant if the variability portion explained by predictor variable on the regression line is grater than the other unexplained portion. Moreover, by using ANOVA Table, we obtain the coefficient of determination R^2, which is defined as the ratio of the sum of squares as a measure of fit of the linear regression model. In other words, the coefficient of determination shows the proportion of variability in y explained by x.

Source	SS	df	MS	F	p-value
Regression	SSR	p	$MSR = \frac{SSR}{p}$	$\frac{MSR}{MSE}$	< 0.05
Error	SSE	$n-p-1$	$MSE = \frac{SSE}{n-p-1}$		
Total	$SSTO$	$n-1$			

Table 3.1: ANOVA table

3.2.2 Residuals Plot

However, the fitted linear regression model might be inadequate for several reasons:

1. The errors(residulas) may not be normally distributed

2. The fitted regression functional form may be incorrect

When processing diagnostic checking, the role of residual plot is also important. Because a residual plot is a graph that shows the residuals ($e = y_i - \hat{y}_i$) on the vertical axis and the fitted value on the horizontal axis, the residuals plot can represent the variation that the regression model has not been able to explain. Therefore, if the fitted model is made, the scatter plot of residuals should resemble a normal distribution. In other words, it should vary in a horizontal area around zero. Any departure from such a horizontal area will be taken as an indication of model inadequacy. Each graph shown below represents an approximate simple regression model, although some deviation from the fitted straight line is observed. Several examples of residuals plots are shown in Figure 3.2.

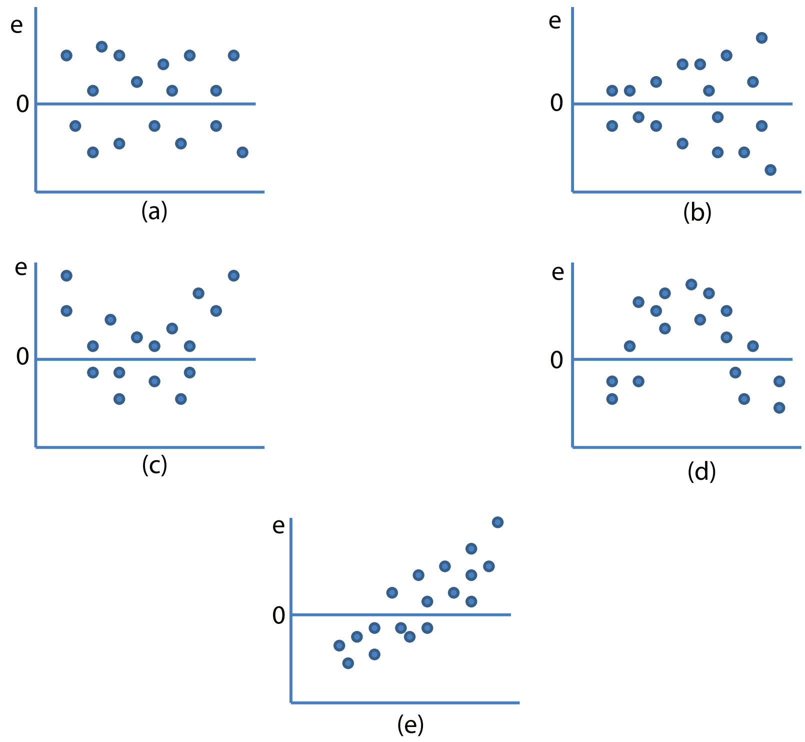

Figure 3.2: Possible residual plots which can be obtained from a linear regression model

Figure 3.2 (a) is a satisfactory plot with the residuals distribution in a horizontal band with no certain pattern. Such a plot indicates an appropriate regression model. The plot of Figure 3.2 (b) represents residuals falling in a funnel shape. Such a plot indicates an increase in variance of residuals because the assumption of constant variance is violated. The residuals following the patterns of Figure 3.2 (c) or (d) that show the linear regression model is not adequate. Addition of higher order terms to the regression model is required in such a case. A plot of residuals can also show a pattern as seen in Figure 3.2 (e) indicating that the residuals increase or decrease as time progresses [3].

3.2.3 $|errors|\,[kWh]$ and $erros$ %

The average monthly $|errors|\,[kWh]$ and the % rate of $error$ can be used as an other criteria to predict a linear regression model accurately. The $|errors|$ and $error$ % are given by

$$|errors| = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{60 \cdot 10^3}\,[kWh] \tag{3.2}$$

$$error\ \% = \frac{error}{P_{\max}} \cdot 100 \tag{3.3}$$

where,

n = the number of observations

y_i = the P_{max} in the ith trial

$\bar{y}$ = mean P_{max}

Chapter 4

Predicting PV Max DC Power

This chapter discusses a detailed evaluation of the accuracy of fitting a linear regression model. The prediction of PV max power was performed based on measured data GH, PA, and I_{sc}. As mentioned before, the two predictor variables, V_{oc} and T_{emp}, are not discussed in this chapter. Due to the fact that V_{oc} and T_{emp} shows large variations over the entire observation, these two variables do not show a linear model adequately.

4.1 Evaluation of a Regression Model

This analysis presents the evaluation of the significance of a linear regression model which is formed by the relationship between P_{max} and other measured data. The evaluation is based on ANOVA table with the null hypothesis, which indicates that fitted regression slope is zero. ($H_0 : \beta_1 = 0$.)

In the linear regression model with radiation measured data, the mean square errors (MSE) of the tilted radiation estimation, PA, was slightly smaller than GH and the other measured PV system data, I_{sc}, showed the smallest MSE in all measured data.

4.1.1 PV MAX DC POWER versus Global Horizontal

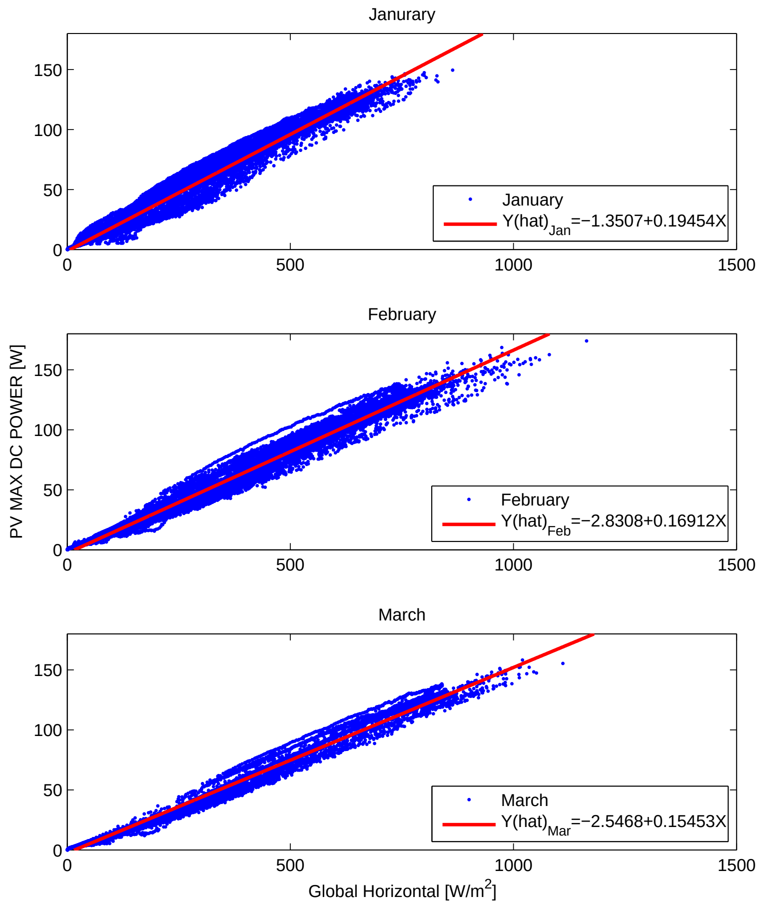

Figure 4.1: Model Fitting of P_{max} vs GH

The plot of the observed data (blue dots for January, green dots for February, and black dots for March) and the fitted straight lines (red lines) are shown in Figure 4.1. The regression models show that each P_{max} increased by 0.19454, 0.16912, and 0.15453 $[W]$ for an increase in a $\left[{}^{W}\!/_{m^2}\right]$ of GH.

Source		SS	df	MS	F	p-value
Regression	Jan.	3.825e+07	1	3.825e+07	8.649e+05	0
	Feb.	3.258e+07	1	3.258e+07	9.544e+05	0
	Mar.	2.242e+07	1	2.242e+07	1.022e+06	0
Error	Jan.	8.894e+05	20110	44.229		
	Feb.	6.792e+05	19898	34.134		
	Mar.	2.875e+05	13102	21.940		
Total	Jan.	3.914e+07	20111			
	Feb.	3.326e+07	19899			
	Mar.	2.271e+07	13103			

Table 4.1: ANOVA of P_{max} vs GH

Table 4.1 shows that each coefficient of determination is given by $R^2 = (3.825e+07)/(3.914e+07) = 0.977$ for January, $R^2 = (3.258e+07)/(3.326e+07) = 0.980$ for February, and $R^2 = (2.242e+07)/(2.271e+07) = 0.987$ for March. This indicates that 97.7%, 98.0%, and 98.7% of the variation is explained by the predictor variable, GH. The F statistic value and p-value indicate the result of $H_1 : \beta_1 \neq 0$ in the test of $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. Therefore, these results indicate that there is enough evidence for a significant positive linear relationship between P_{max} and GH.

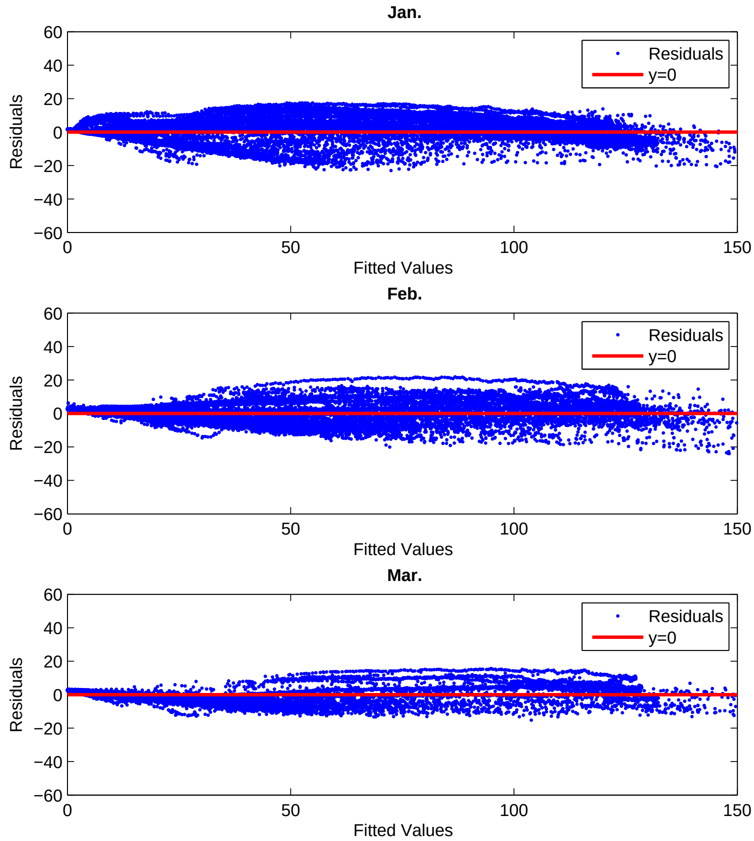

Figure 4.2: Residuals Plot of P_{max} vs GH

The residuals plot in Figure 4.2 shows that the fitted model is reasonable. The residuals vary in a horizontal area around zero ($Y = 0$), in other words, the residuals have a normal distribution on the horizontal axis; consequently, these residuals accurately represent the variation.

4.1.2 PV MAX DC POWER versus Plane of Array

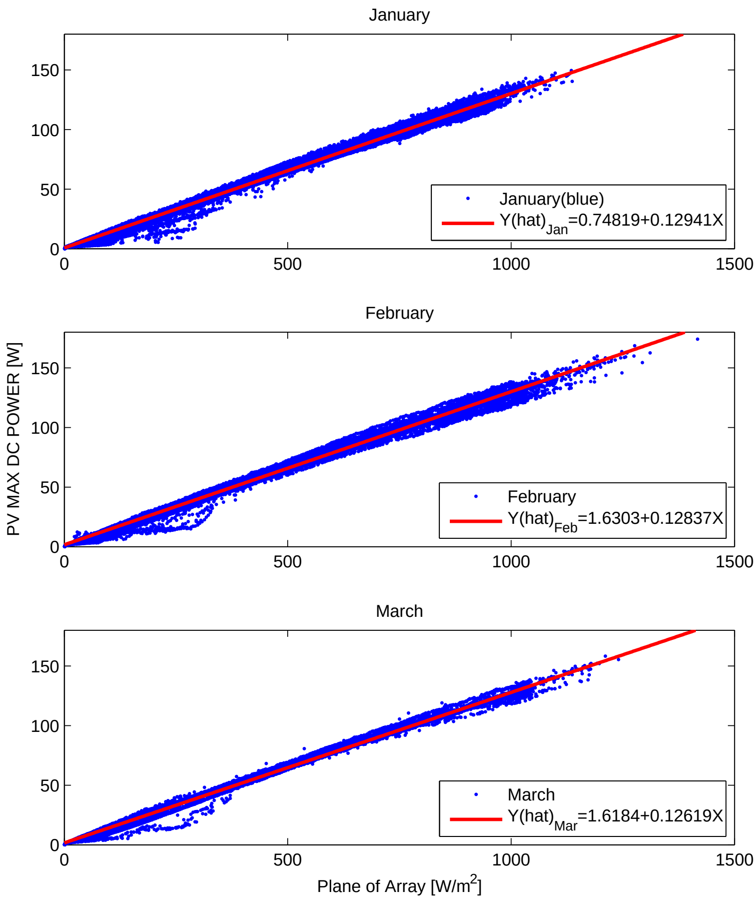

Figure 4.3: Model Fitting of P_{max} vs PA

The fitted regression models as shown above in Figure 4.3 suggest that each P_{max} is increasing by increasing PA. All associations of P_{max} and PA are linear. These regression model functions are almost identical as $\hat{Y}_{Jan} = 0.74819 + 0.12941X$, $\hat{Y}_{Feb} = 1.6303 + 0.12837X$, and $\hat{Y}_{Mar} = 1.6184 + 0.12619X$.

Source		SS	df	MS	F	p-value
Regression	Jan.	3.893e+07	1	3.893e+07	3.756e+06	0
	Feb.	3.306e+07	1	3.306e+07	3.266e+06	0
	Mar.	2.262e+07	1	2.262e+07	3.588e+06	0
Error	Jan.	2.085e+05	20110	10.367		
	Feb.	2.014e+05	19898	10.121		
	Mar.	8.261e+04	13102	6.305		
Total	Jan.	3.914e+07	20111			
	Feb.	3.326e+07	19899			
	Mar.	2.271e+07	13103			

Table 4.2: ANOVA of P_{max} vs PA

Table 4.2 shows that the coefficient of determination is given as monthly order $R^2_{Jan} = (3.893e+07)/(3.914e+07) = 0.995$, $R^2_{Feb} = (3.306e+07)/(3.326e+07) = 0.994$, and $R^2_{Mar} = (2.262e+07)/(2.271e+07) = 0.996$. The results indicates that 99.5%, 99.4, and 99.6 of the variation are explained by the PA variable. All p-value, listed in Table 4.2 are 0. These refer to the alternative hypothesis ($H_1 : \beta_1 \neq 0$). Therefore, P_{max} and PA have a significant positive linear relationship.

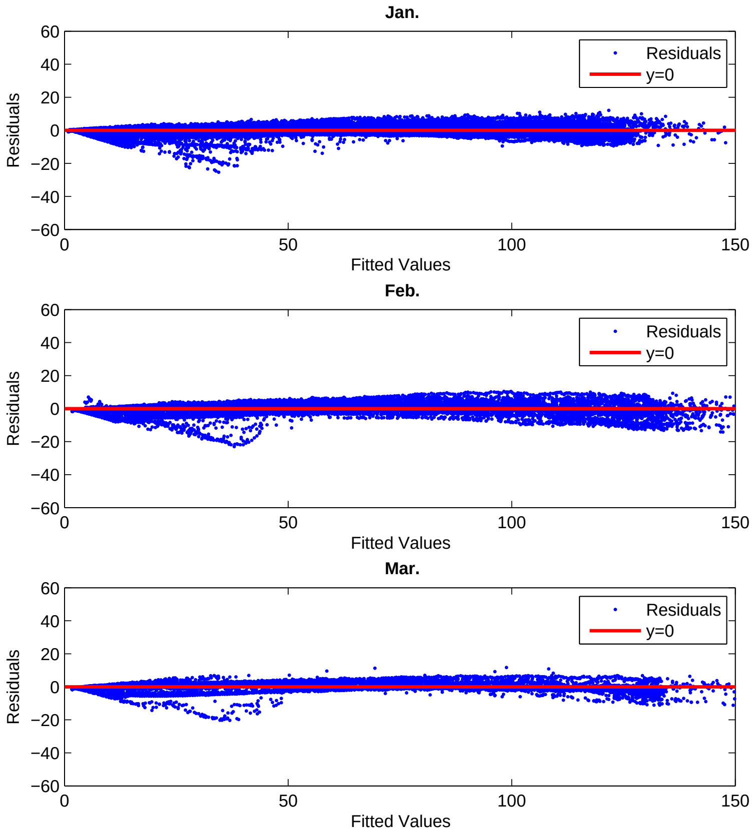

Figure 4.4: Residuals Plot of P_{max} vs PA

Figure 4.4 represents the residuals plot of PA. Since the residuals of three months vary in a horizontal area around zero ($Y = 0$), these residuals plot suggest that the fitted models are adequate.

4.1.3 PV MAX DC POWER versus Short Circuit Current

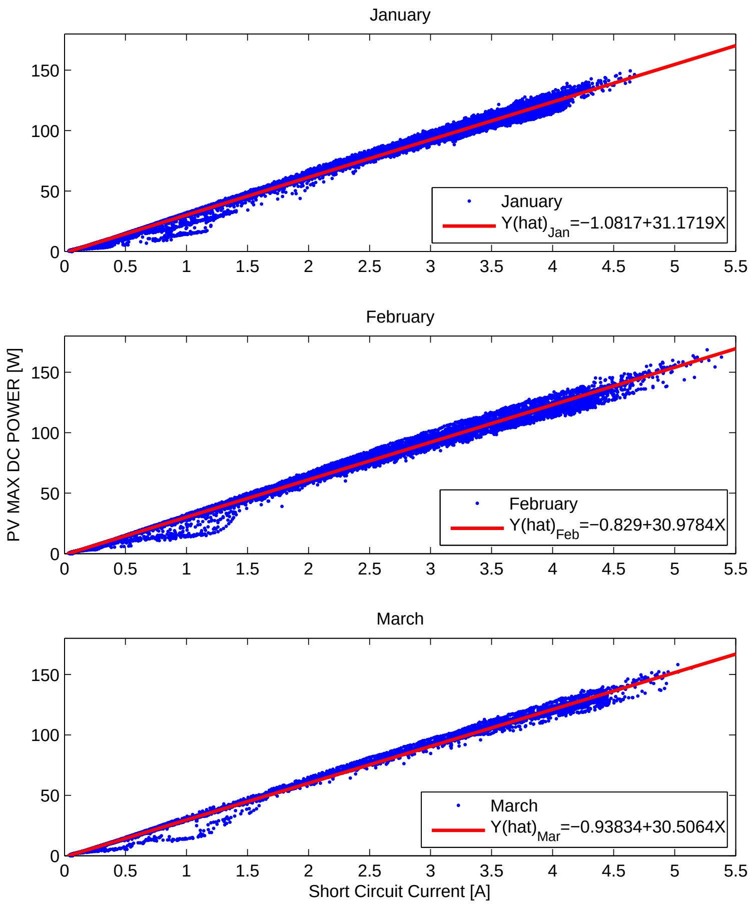

Figure 4.5: Model Fitting of P_{max} vs I_{sc}

Figure 4.5 shows that the distributions of all observed data are almost identical to the fitted straight lines. These regression models ($\hat{Y}_{Jan} = -1.0817 + 31.1719X$, $\hat{Y}_{Feb} = -0.829+30.9784X$, and $\hat{Y}_{Mar} = -0.93834+30.5064X$) have a strong positive linear relation between the response and the predictor variable.

Source		SS	df	MS	F	p-value
	Jan.	3.899e+07	1	3.899e+07	4.992e+06	0
Regression	Feb.	3.310e+07	1	3.310e+07	4.264e+06	0
	Mar.	2.265e+07	1	2.265e+07	5.503e+06	0
	Jan.	1.570e+05	20110	7.809		
Error	Feb.	1.545e+05	19898	7.763		
	Mar.	5.394e+04	13102	4.117		
	Jan.	3.914e+07	20111			
Total	Feb.	3.326e+07	19899			
	Mar.	2.271e+07	13103			

Table 4.3: ANOVA of P_{max} vs I_{sc}

Table 4.3 shows all coefficient of determination which are calculated by $R^2_{Jan} = (3.899e+07)/(3.914e+07) = 0.996$, $R^2_{Feb} = (3.310e+07)/(3.326e+07) = 0.995$, and $R^2_{Mar} = (2.265e+07)/(2.271e+07) = 0.997$. These indicates that all variations explained by the I_{sc} show high values of 99.6%, 99.5%, and 99.7%. The F test, example for January, in the ANOVA table for the I_{sc} predictor variable is equal to $(3.899e{+}07)/(7.809) = 4.992e{+}06$. The distribution is F(1, 20110), and the probability of observing a value greater than or equal to 4.992e+06 is less than $\alpha = 0.05$. There is strong evidence that β_1 is not equal to zero. Therefore, P_{max} and I_{sc} have a significant positive relationship. (Feb. and Mar. are applied as a same way)

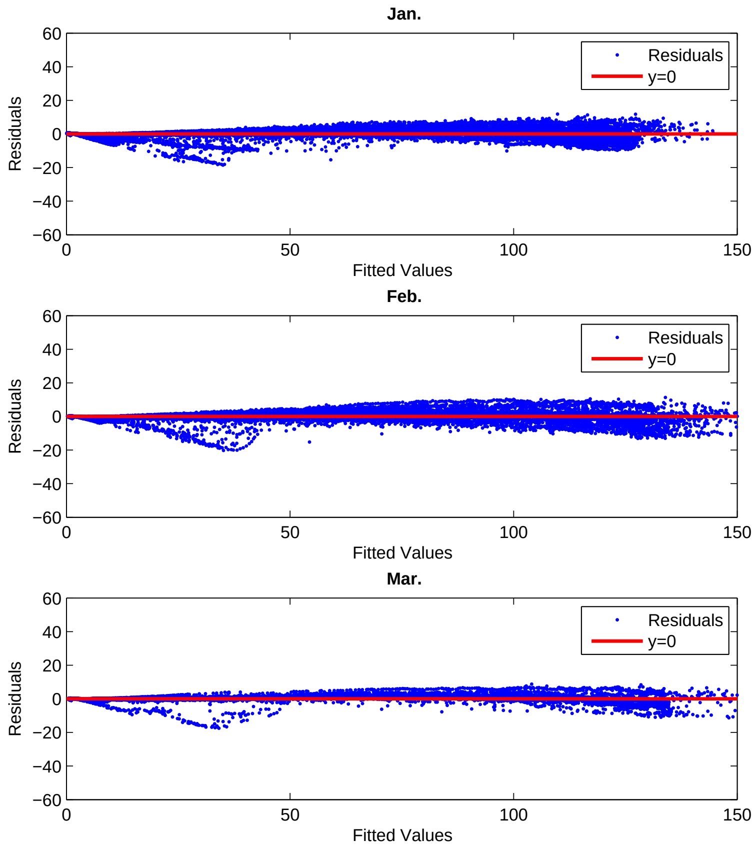

Figure 4.6: Residuals Plot of P_{max} vs I_{sc}

Figure 4.6 shows that three fitted model are adequate because the residuals vary in a horizontal area around zero.

4.1.4 PV MAX DC POWER versus PV Panel Temperature

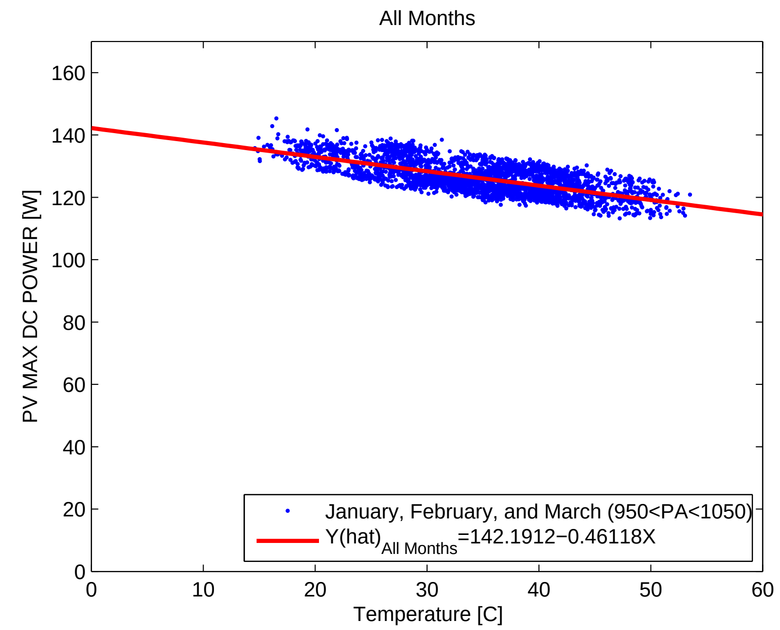

Figure 4.7: Model Fitting of P_{max} vs T_{emp}

Figure 4.7 shows that P_{max} and T_{emp} have a negative correlation. All 3 months data points are in the range of PA=950 to PA=1050. From the figure above, it is clear that P_{max} decreases as T_{emp} increases at the rate of 0.461 $[W/°C]$ for a 130 $[W]$ panel pair. The result corresponds to 0.35 % per $[°C]$.

4.2 Predicting PV MAX Power

The following figures show a comparison between the predicted P_{max} features and the measured P_{max}. The predicted P_{max} is obtained from the fitted linear regression model. As shown above, in the ANOVA table, the smaller the MSE value, the better the estimated curve is.

4.2.1 The results of Predicting PV MAX Power Plots

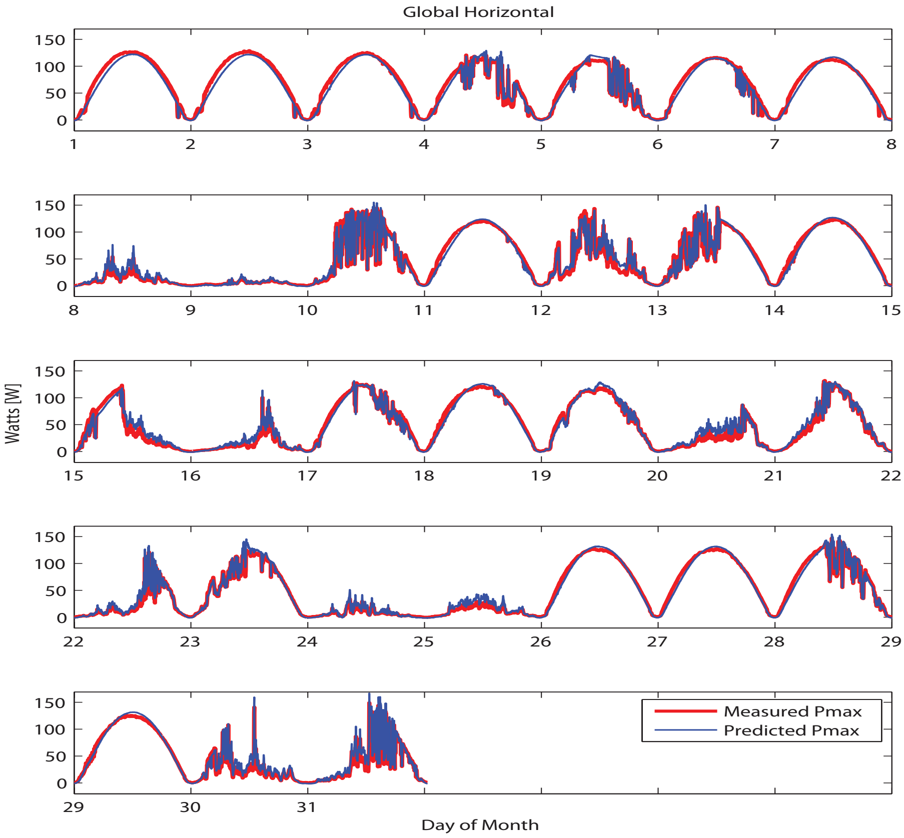

Figure 4.8: Predicted Max PV Power using GH, 130$[W]$ Test Panel, January, 2012

	$P_{\max}$ $[kWh]$	$\lvert Error \rvert$ $[kWh]$	$Error\ Rate$ %
January	17.896	1.695	9.472

Table 4.4: Summary of Predicting P_{max} using GH

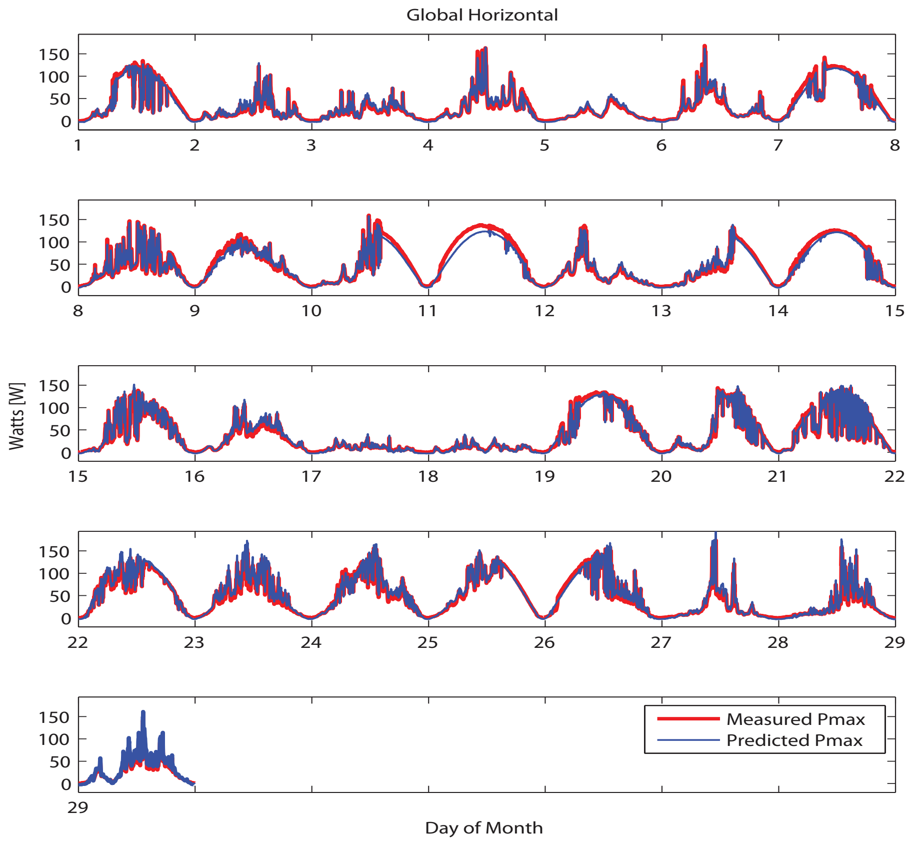

Figure 4.9: Predicted Max PV Power using GH, 130$[W]$ Test Panel, February, 2012

	$P_{\max}$ $[kWh]$	$\lvert Error \rvert$ $[kWh]$	$Error\ Rate$ %
February	14.409	1.452	10.08

Table 4.5: Summary of Predicting P_{max} using GH

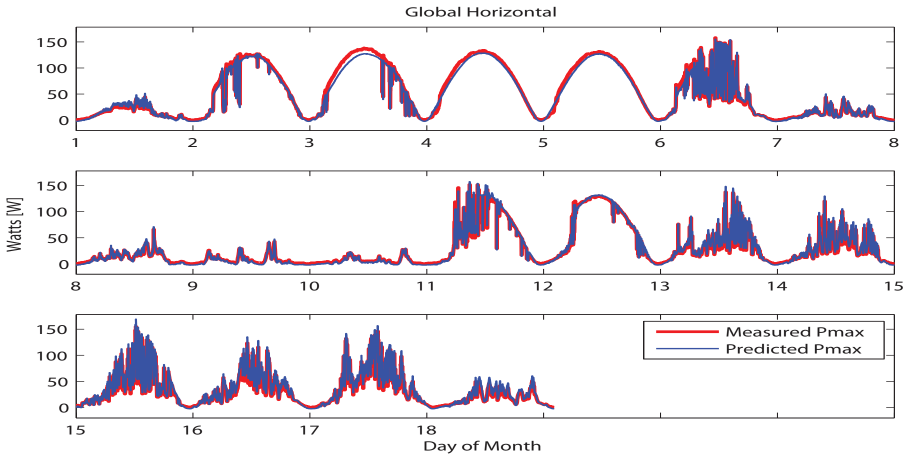

Figure 4.10: Predicted Max PV Power using GH, 130$[W]$ Test Panel, March, 2012

	$P_{\max}$ $[kWh]$	$\lvert Error \rvert$ $[kWh]$	*Error Rate* %
March	8.551	0.795	9.292

Table 4.6: Summary of Predicting P_{max} using GH

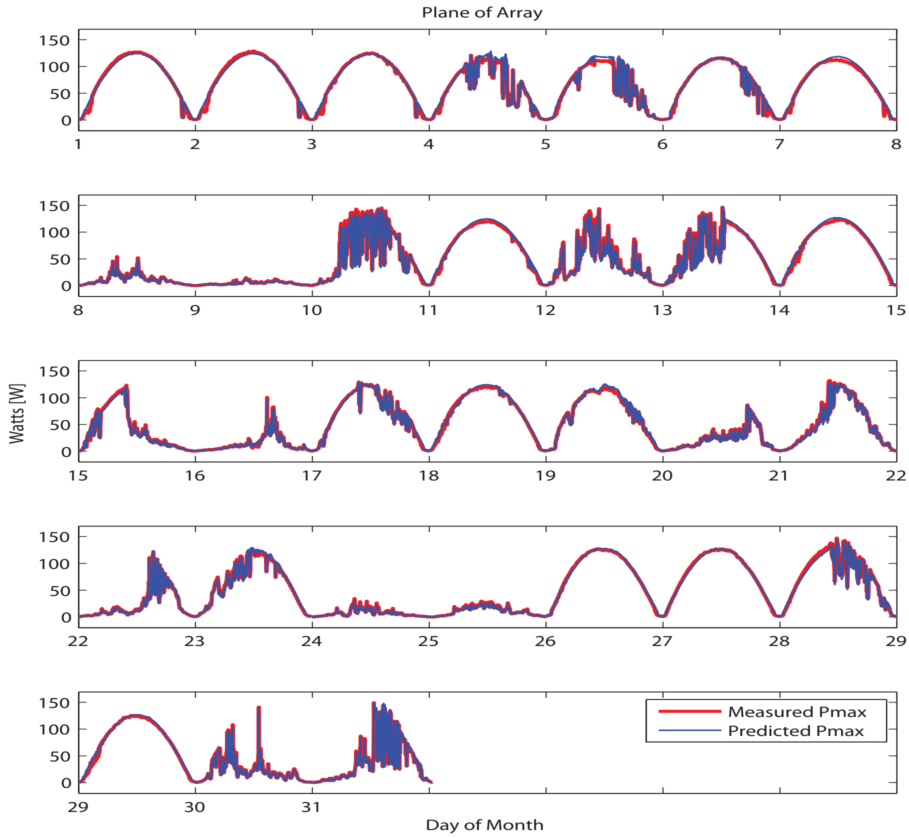

Figure 4.11: Predicted Max PV Power using PA, 130$[W]$ Test Panel, January, 2012

	$P_{\max}$ $[kWh]$	$\|Error\|$ $[kWh]$	$Error\ Rate$ %
January	17.896	1.010	5.643

Table 4.7: Summary of Predicting P_{max} using PA

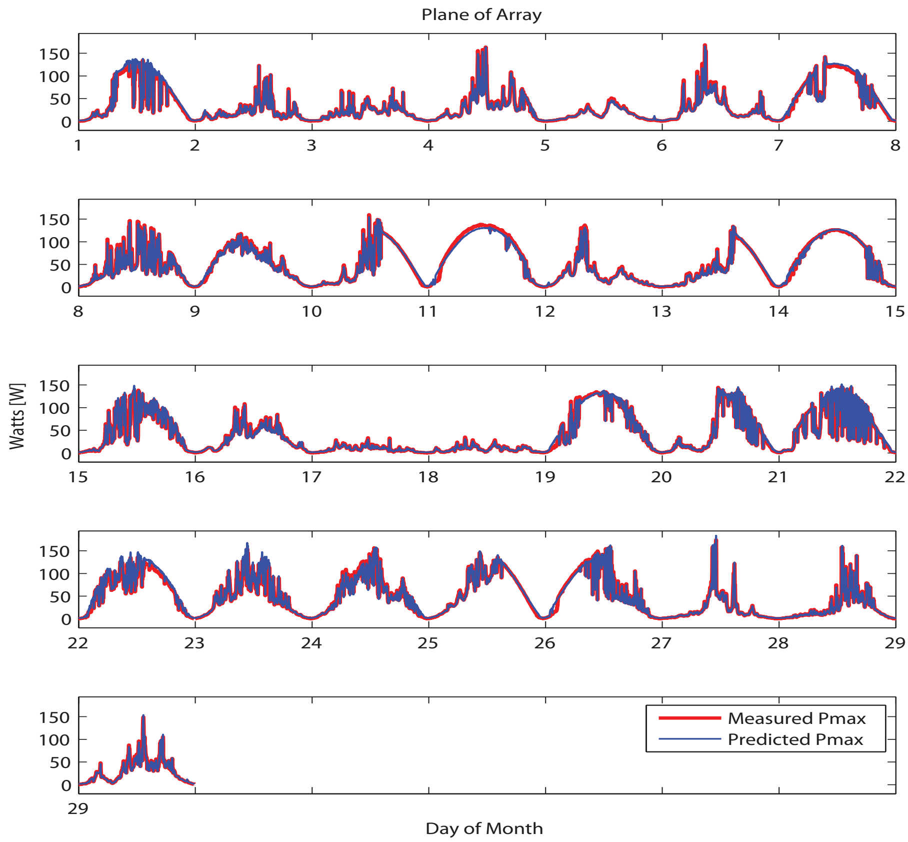

Figure 4.12: Predicted Max PV Power using PA, 130[W] Test Panel, February, 2012

	$P_{\max}$ [kWh]	$\lvert Error \rvert$ [kWh]	$Error\ Rate$ %
February	14.409	1.102	7.651

Table 4.8: Summary of Predicting P_{max} using PA

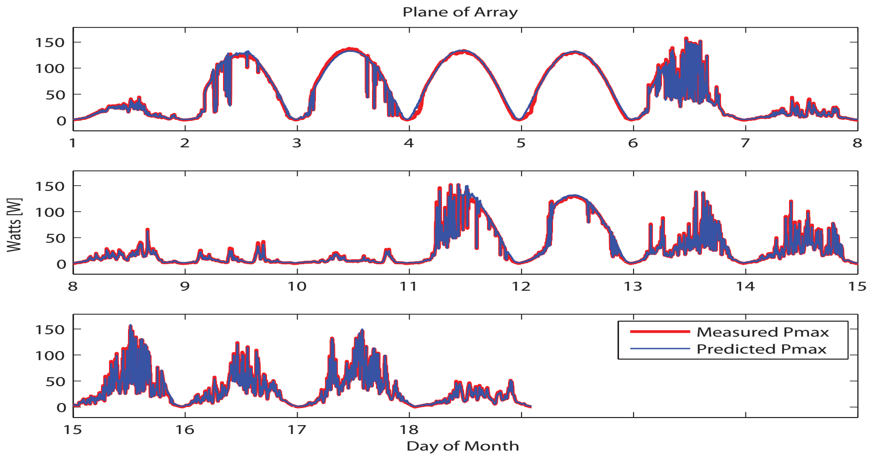

Figure 4.13: Predicted Max PV Power using PA, 130[W] Test Panel, March, 2012

	$P_{\max}$ [kWh]	$\lvert Error\rvert$ [kWh]	$Error\ Rate$ %
March	8.551	0.670	7.832

Table 4.9: Summary of Predicting P_{max} using PA

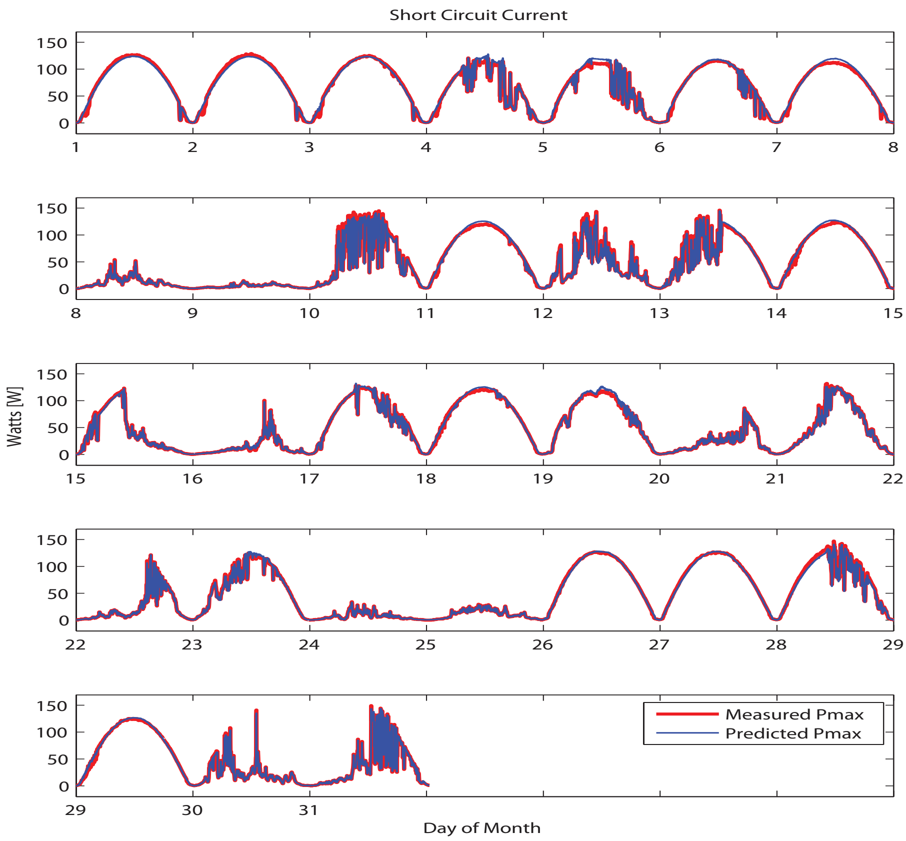

Figure 4.14: Predicted Max PV Power using I_{sc}, 130$[W]$ Test Panel, January, 2012

	$P_{\max}$ $[kWh]$	$\lvert Error \rvert$ $[kWh]$	$Error\ Rate$ %
January	17.896	0.586	3.277

Table 4.10: Summary of Predicting P_{max} using I_{sc}

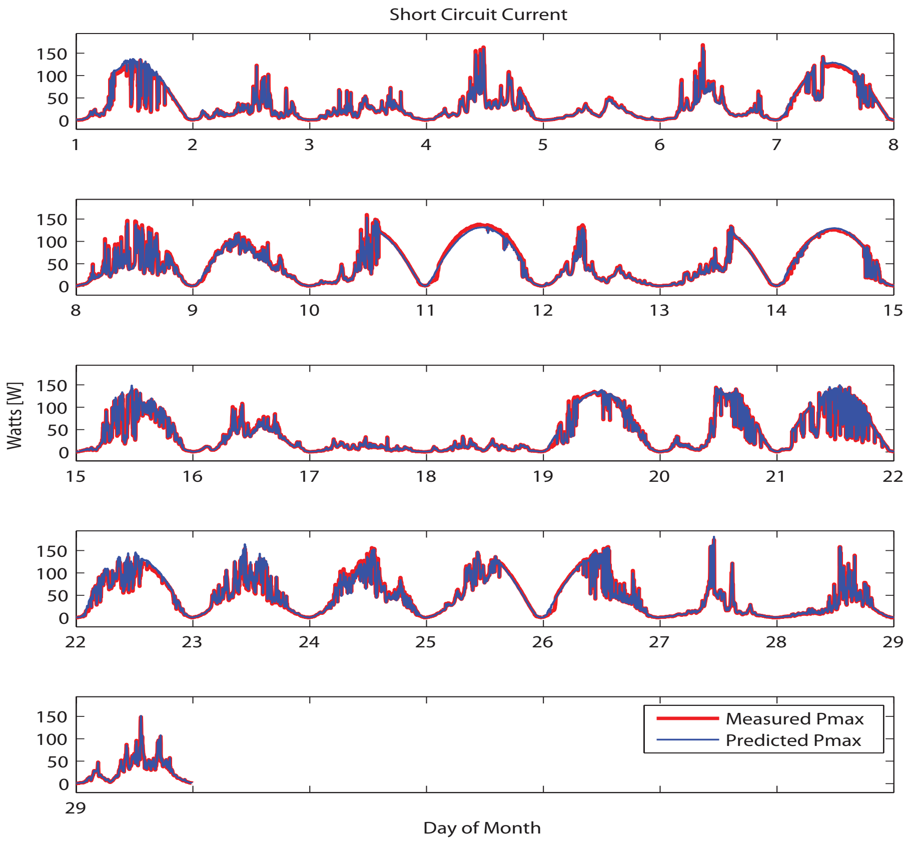

Figure 4.15: Predicted Max PV Power using I_{sc}, $130[W]$ Test Panel, February, 2012

	$P_{\max}$ $[kWh]$	$\lvert Error \rvert$ $[kWh]$	$Error\ Rate$ %
February	14.409	0.545	3.783

Table 4.11: Summary of Predicting P_{max} using I_{sc}

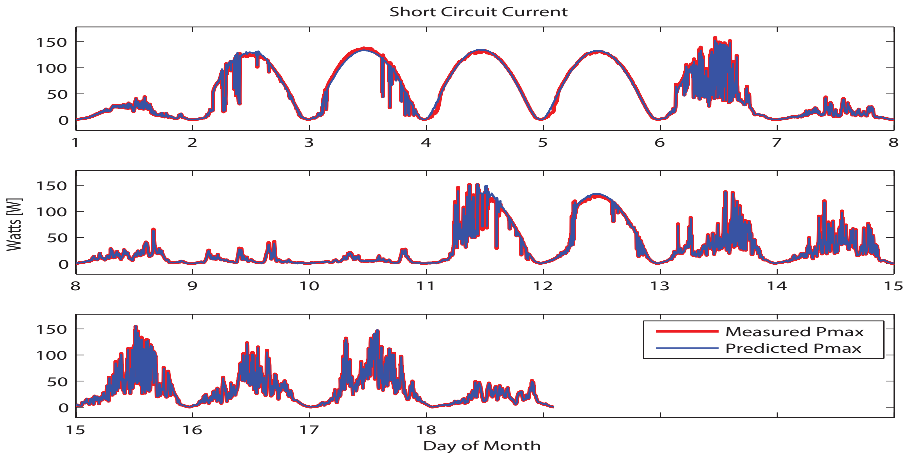

Figure 4.16: Predicted Max PV Power using I_{sc}, 130[W] Test Panel, March, 2012

	$P_{\max}$ [kWh]	$\lvert Error \rvert$ [kWh]	$Error\ Rate$ %
March	8.551	0.232	2.713

Table 4.12: Summary of Predicting P_{max} using I_{sc}

Chapter 5

Conclusion

This study presents the results of predicting PV max dc power using a linear regression model based on measured radiation and panel data. The results revealed that although there are some variations from the regression line, the correlation of each response variable and predictor variable is highly linear. The regression models using GH, PA, and I_{sc} could accurately predict the max solar power from the I-V curve tracer. Specifically, in measured radiation data, the tilted radiation data, PA, shows lower errors as a monthly order (MSE of 10.367, 10.121, and 6.305, $|Errors|\,[kWh]$ of 1.010, 1.102, and 0.670) than GH errors (MSE of 44.229, 34.134, and 21.940, $|Errors|\,[kWh]$ of 1.695, 1.452, and 0.795). In measured PV panel data, an optimum predictor variable can be achieved by I_{sc} with the smallest errors (MSE of 7.809, 7.763, and 4.117, $|Errors|\,[kWh]$ of 0.586, 0.545, and 0.232) which show the accuracy of the prediction of PV max dc power. Although using two predictor variables, PA and I_{sc}, shows more precisely predicted values, the most widely used predictor variable, GH, also can help explain variation statistically in the regression model and predict the max solar power.

Furthermore, the monthly fitted linear regression was investigated to accurately predict an average monthly PV max power through a regression coefficient trend. The monthly order slope of the linear regression line from January to March slightly decreased because the distribution of P_{max} changed, which indicates that the angle between the sun and the equator changes each month.

However, further study of the predicting regression model is recommended to expanded with sufficiently long term data (e.g., season or year), which can statistically predict PV max dc power more precisely.

Bibliography

[1] N. H. Bingham and John M. Fry. "Regression Linear Models in Statistics," Springer, 2010.

[2] Alan Agresti and Barbara Finlay, "Statistical Methods for the Social Sciences Fourth Edition," Prentice Hall Inc, 2009.

[3] Gukbert M. Masters, "Renewable and Efficient Electric Power Systems", John Wiley and Sons, Inc, 2004.

[4] W. Mack Grady, Class Lecture, Department of Electrical and Computer Engineering, EE462L: Power Electronics, Solar Power, I-V characterisitc, the University of Texas at Austin, Fall 2010.

[5] B. Abraham and J. Ledolter, "Statistical Methods for Forecasting," John Wiley and Sons Inc, 1983.